UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO

CARRERA:
INGENIERÍA ELECTRÓNICA

Trabajo de titulación previo a la obtención del título de:
INGENIEROS ELECTRÓNICOS

TEMA:
ASIGNACIÓN DINÁMICA DE RECURSOS EN REDES VANET MEDIANTE APRENDIZAJE POR REFUERZO

AUTORES:
OSCAR EDUARDO CAIZA CHAFLA
CHRISTIAN ALEXANDER JAMI HERRERA

TUTOR:
JOSÉ RENATO CUMBAL SIMBA

Quito, mayo del 2021

Cesión de derechos de autor

Nosotros, Oscar Eduardo Caiza Chafla y Christian Alexander Jami Herrera, con documentos de identificación N° 1725512345, N° 1718493495 respectivamente, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación intitulado ASIGNACIÓN DINÁMICA DE RECURSOS EN REDES VANET MEDIANTE APRENDIZAJE POR REFUERZO, mismo que ha sido desarrollado para optar por el título de: Ingenieros Electrónicos, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

....

Oscar Eduardo Caiza Chafla
C.C: 1725512345

Christian Alexander Jami Herrera
C.C: 1718493495

Quito, mayo del 2021

Declaratoria de coautoría del docente tutor

Yo declaro que bajo mi dirección y asesoría fue desarrollado el trabajo de titulación ASIGNACIÓN DINÁMICA DE RECURSOS EN REDES VANET MEDIANTE APRENDIZAJE POR REFUERZO realizado por Oscar Eduardo Caiza Chafla y Christian Alexander Jami Herrera, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana para ser considerados como trabajo final de titulación.

Quito, mayo de 2021

José Renato Cumbal Simba
C.C: 1714663760

ASIGNACIÓN DINÁMICA DE RECURSOS EN REDES VANET MEDIANTE APRENDIZAJE POR REFUERZO

Renato Cumbal§, Oscar Caiza†, Cristian Jami‡
Universidad Politécnica Salesiana Ecuador
§Email: jcumbal@ups.edu.ec
†Email: ocaizac@est.ups.edu.ec
‡Email: cjamih@est.ups.edue.ec

***Resumen*— En el presente trabajo se propone el despliegue de la infraestructura RSU a través de un algoritmo de aprendizaje por refuerzo para distribuir de una manera óptima los recursos en la red vehicular. El objetivo principal de nuestro estudio es utilizar el algoritmo Q-Learning para asignar canales desde un controlador hacia las RSU en el escenario de planeación. Con este despliegue inicial y su movilidad se realizará un análisis a través de un modelo de optimización para obtener una cantidad mínima de dispositivos en la infraestructura de la VANET simulada. El aprendizaje del algoritmo sobre el escenario se establece dinámicamente con relación a la demanda vehicular y sus restricciones de cobertura para una comunicación V2I.**

Palabras Clave— Infraestructura Vanet; RSU; optimización; aprendizaje por refuerzo; V2I.

***Abstract*— In this paper we propose the deployment of the RSU infrastructure through a reinforcement learning algorithm to optimally distribute the resources in the vehicular network. The main objective of our study is to use the Q-Learning algorithm to allocate channels from a controller to the RSUs in the planning scenario. With this initial deployment and its mobility, an analysis will be performed through an optimization model to obtain a minimum number of devices in the simulated VANET infrastructure. The learning of the algorithm on the scenario is dynamically established in relation to the vehicular demand and its coverage restrictions for a V2I communication.**

Keywords— Vanet Infraestructure; RSU; optimization; reinforcement learning; V2I.

I. INTRODUCCIÓN

Las redes ad-hoc vehiculares (VANET) en la actualidad se han convertido en uno de los campos de investigación más llamativos de las comunicaciones inalámbricas en los sistemas inteligentes de transporte (ITS), mismos que buscan mejorar la movilidad con una adecuada gestión del tráfico vehicular, la reducción de la tasa de accidentes en las vías y manejo del consumo energético así como la prestación de algunos servicios tales como: acceso a internet, seguridad vial, reportes climatológicos, noticias en tiempo real, entre otros [1].

En la figura 1 se puede apreciar la arquitectura de las redes VANET la cual está conformada por nodos móviles que son vehículos con sensores incorporados llamados OBU (On-Board Unit) que su función es la recepción y transmisión de datos; y por nodos fijos o antenas que son unidades a lado del camino llamados RSU (Road-Site Unit) los cuales se encargan de compartir la información y funcionan como la puerta de enlace para los nodos móviles [2].

Para proporcionar los servicios antes mencionados en esta red existen tres modalidades de comunicación, la primera es una comunicación entre nodos móviles, es decir de vehículo a vehículo (V2V), la segunda comunicación posible es la que se efectúa entre el nodo móvil y la puerta de enlace, es decir entre el vehículo y la RSU (V2I). Con el pasar de los años estas comunicaciones se han ido mejorando y con el desarrollo de 5G y el internet de las cosas (IoT) se logra otra comunicación la cual se realiza entre el nodo móvil y cualquier elemento de la red VANET, de tal manera, se habla de una comunicación de vehículo con todo (V2X) [3].

Uno de los principales retos de las redes VANET es la correcta ubicación de la infraestructura (RSU) para que los vehículos puedan comunicarse con las mismas, por lo tanto, en el artículo se busca minimizar el número de antenas necesarias en un espacio determinado de estudio o escenario en función del radio de cobertura y la capacidad de las RSU. [4] Para optimizar la cantidad de RSU se emplea un modelo de programación lineal entera (ILP), en el cual se considera un número de antenas específico para cubrir el mayor territorio posible y el mayor porcentaje de usuarios (OBU´s) en el escenario de estudio [5].

Debido al frecuente cambio de la topología, la conexión intermitente de los terminales y la movilidad vehicular aleatoria en las vías, las redes VANET se enfrentan a una dificultad en su funcionamiento [6], por esto se han propuesto varias alternativas en búsqueda de una asignación óptima de recursos [7]. Hace algunos años estas propuestas únicamente se basaban en modelos matemáticos con un alto coste computacional lo cual hace que el rendimiento de la red no sea el esperado como se puede ver en [8]. Hoy en día

gracias al desarrollo tecnológico se puede pensar en una alternativa basada en la inteligencia artificial (IA) [9].
El aprendizaje por refuerzo (RL) [10][11] es una categoría del aprendizaje automático, la cual es una herramienta eficaz y muy recomendada para abordar procesos de decisión de Markov (MDP) [12]; el RL permite a un agente aprender la política óptima a través de la interacción con un entorno mediante prueba y error utilizando la retroalimentación de sus propias acciones y experiencias según las recompensas recibidas. Q-Learning es el método más eficiente y ampliamente recomendado por la literatura e investigaciones como se menciona en [13][14].

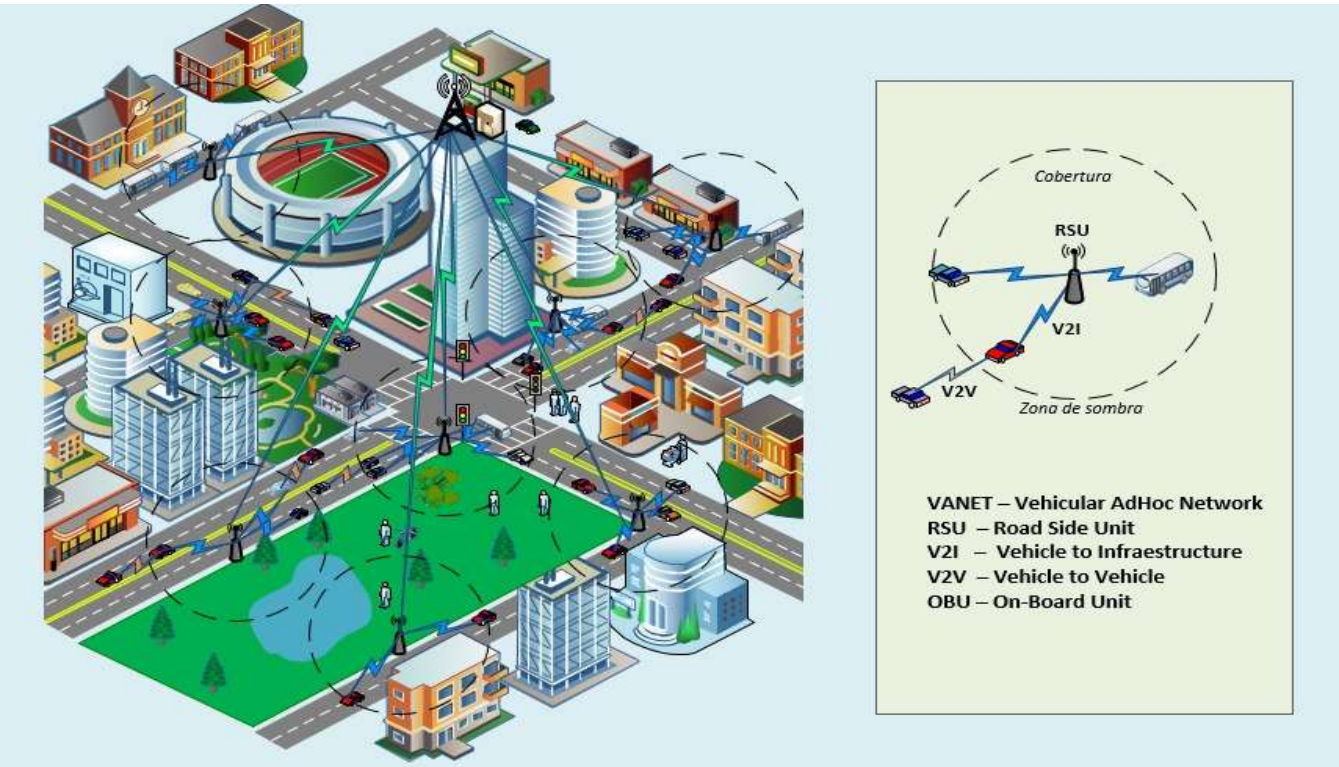

Figura. 1. Arquitectura VANET

La principal diferencia entre otros algoritmos de IA es su proceso de 4 pasos que son los siguientes: El primero son las políticas, que relacionan los estados con aquellas acciones que va a tomar el agente; el segundo paso se refiere a la función de recompensa, que es la encargada de realizar el análisis del estado actual y asignar una recompensa o penalización según los resultados de la acción tomada; el tercer paso es la función de valor, en esta etapa se predice la recompensa futura que se le dará al agente para una acción en el futuro; y finalmente el cuarto paso se refiere al modelo de entorno, que se encarga de hallar los estados y las acciones posibles que se pueden tomar por el agente de manera inteligente [15].

El enfoque Q-Learning (QL) es un tipo de algoritmo de RL donde el agente crea una matriz Q, la cual será su memoria, que la obtiene después de interactuar con el entorno [16]. Al explorar el entorno, la función Q actúa como función de valor, estimando recompensas futuras ante acciones actuales. Los Q valores se almacenan en una tabla para cada estado y cada acción posible. Sin embargo, la matriz puede ser bastante compleja e inaplicable para grandes espacios de acción de estado [17].

El resto de este documento se organiza de la siguiente manera. En la sección II se presenta desde una perspectiva general trabajos similares desde la visión de otros autores. En la sección III se introduce el planteamiento del problema, la función objetivo, las restricciones, el modelo de optimización y el algoritmo Q-Learning utilizado en la asignación dinámica de recursos en la red VANET, para luego en la sección IV mostrar los resultados obtenidos y el análisis de estos al aplicar los algoritmos en el escenario de planeación. Finalizamos este artículo con la sección V relacionada con las conclusiones y trabajos futuros de estudio.

II. Asignación dinámica de recursos en una red VANET

A. Optimización de infraestructura en VANET

Al simular una red vehicular se puede modelar diferentes escenarios de movilidad para un correcto estudio como se menciona en [18] existe la herramienta de red de proyección para sistemas ITS llamada Simulación para Movilidad Urbana (SUMO), el cual es un simulador de código abierto, que proporciona usabilidad y funciones sin restricciones para desarrollar espacios de tráfico discretos capaces de imitar con un enfoque realista el comportamiento de los vehículos; SUMO puede importar mapas con carreteras e información sobre las señales de tráfico y orientación de las vías de Open Street Map. Como se puede observar en la figura 2 se han generado los escenarios de estudio para el presente documento.

En [19] se aprecia el avance de los últimos años en cuanto a diferentes estándares para asegurar la interoperabilidad e intercambio de información de los vehículos, los estándares se basan en el IEEE 802.11p el cual establece la comunicación en las VANETs, pero se enfrentan a grandes retos como el corto tiempo de conexión de los nodos móviles

a la infraestructura (RSU) en las comunicaciones V2I y con esto la correcta asignación de los recursos a los vehículos que están dentro de la red.

Figura 2: Escenario de estudio

El canal de propagación en estos casos presenta características especiales y es el factor más importante comparado con otro tipo de sistemas inalámbricos, por ejemplo, la rápida variación en tiempo y el dinamismo de las estadísticas del canal por el tipo de entorno de movilidad vehicular en las vías, por lo tanto, es sumamente importante el manejo de este recurso mediante un modelo matemático dinámico como se estable en [20].

En [21] se asegura que se debe considerar en las redes VANET un despliegue óptimo de la infraestructura y esto representa un reto, el cual tiene una relación directa con la calidad del manejo de recursos de radio (RRM), en las comunicaciones V2V la calidad depende de la cantidad de vehículos así como en V2I dependen de la cantidad y la ubicación de las antenas (RSU), evidentemente para el último tipo de conexión en ambientes rurales el número de RSU´s no es tan importante, puesto que la densidad del flujo vehicular no es la misma de los ambientes urbanos, que son el tipo de escenarios que se estudian en el presente documento. Por esta razón se debe tener una planificación adecuada para tener un óptimo funcionamiento de la red VANET, es decir usar la menor cantidad posible de RSU´s para cubrir la mayor cantidad de territorio y usuarios. Para la solución de este reto se plantearon varias alternativas basadas en la programación lineal entera (ILP) y métodos heurísticos.

B. Asignación Dinámica De Recursos Mediante Aprendizaje Por Refuerzo En Una Red Vanet

Los autores en [22] mencionan que el aprendizaje automático (ML) son técnicas con las cuales un programa puede aprender a realizar una tarea específica en función a la experiencia que este algoritmo vaya ganando de su entorno; se habla de cuatro categorías como se puede apreciar en la figura 3 son: aprendizaje supervisado, no supervisado, por refuerzo (RL) y aprendizaje profundo (DL). En este trabajo se aplicará RL, donde un agente puede mejorar una tarea asignada mientras va interactuando con su ambiente. También se define que un agente puede ser algo o alguien que descubre y ejecuta, por lo tanto, en este documento se define al agente como un controlador de red.

Figura 3: Ramas principales de Machine Learning

La técnica de aprendizaje por refuerzo tiene un enfoque computacional para comprender y automatizar como tomar las mejores decisiones (políticas) con la finalidad de alcanzar los objetivos necesarios; este tipo de técnica se distingue de las demás, porque el aprendizaje lo realiza un agente mediante la interacción con su entorno, sin la necesidad de alguna supervisión o conocimiento previo del ambiente, es una de las principales ventajas como lo señalan en [23]; de igual manera se recalca que RL denota la interacción que tiene un agente de aprendizaje y su entorno mediante estados, recompensas, acciones y de esta manera se busca la representación sencilla de un problema de inteligencia artificial. También se afirma que el aprendizaje por refuerzo es el que se utiliza por naturaleza, porque todo ser vivo tiene un ambiente de interacción del cual obtiene información como resultado de prueba y error que realiza, es decir las acciones con las cuales busca generar una estrategia para alcanzar una tarea. Es necesario indicar ahora los diferentes componentes de esta técnica.

En [15] se reconoce también que un agente aprende los mejores comportamientos o los óptimos, ya que participa de un entorno mediante prueba y error únicamente con la finalidad de obtener la recompensa más alta. Además, mencionan que el entorno de interacción es modelado con procesos de decisión de Markov (MDP). En [24] se establece que el modelo formal para estos problemas son MDP, los cuales constan de 4 partes fundamentales que son: el espacio de estados (S), la función que contiene las posibles acciones para cada estado (A), la función de transición (T) y la función de recompensa (R), en donde las funciones de transición y recompensa serán determinadas por la función de estados (estado actual), y por la función de acciones.

Todo este proceso tiene como objetivo hallar un conjunto de acciones para tener como resultado la mayor recompensa posible, a las cuales se las conoce como política (G), la cual representa como un agente reacciona en un tiempo t a un determinado estado como lo mencionan en [25]. RL está definido matemáticamente por la expresión de la ecuación 1 que se muestra a continuación:

$$G_t = \sum_{i=1}^{\infty} \gamma^{i-1} R_{t+i} = R_{t+1} + \gamma G_{t+1} \qquad (1)$$

En donde γ es el factor de descuento y R_t representa a la función de recompensa evaluada en cada tiempo t. [15] En la figura 4 se muestra el modelo de RL.

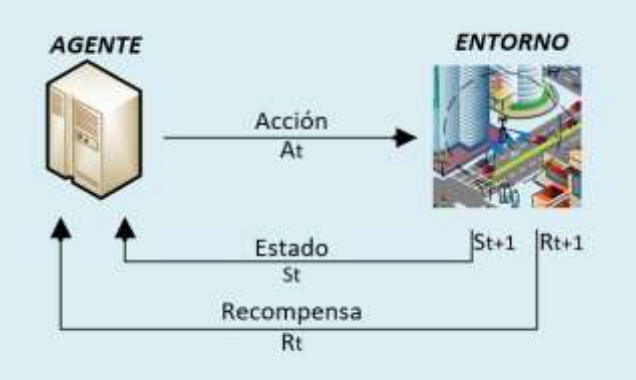

Figura 4: Modelo de aprendizaje por refuerzo

Q-Learning es uno de los tantos algoritmos de RL, el cual presenta varias ventajas como el hecho de ser un modelo libre y que está basado en MDP como modelo matemático y con esto se logra una precisa definición de la tarea que se desea resolver; de los métodos de aprendizaje por refuerzo en [26] se afirma que QL es el más simple de implementar y comprender, además se menciona que otro aspecto significativo es que no requiere conocimiento previo del entorno, lo cual es de gran utilidad para las redes ad-hoc vehiculares, en estos entornos el nivel de incertidumbre es alto debido a que sus variables son aleatorias. En [27] se realiza una demostración de convergencia a la política óptima. Sin embargo, la principal deficiencia de este método es su lentitud, esto quiere decir que el agente requiere prolongados tiempos de aprendizaje para llegar a una política aceptable. En [13] se establece que el componente más importante de este algoritmo es una correcta estimación del valor Q, el cual puede ser hallado mediante una función de aproximación que en algunos casos puede ser no lineal o por una simple tabla de búsqueda.

En [15] se dice que un enfoque clásico para resolver problemas de RL es la búsqueda de la matriz Q y para esto se plantea el uso de una función de estimación $Q(s,a)$, donde $\forall\, s \in S$ y $\forall\, a \in A$ la cual estima la sumatoria de recompensas al tomar una acción a en un determinado estado s. Se dice también que la función Q óptima es la máxima recompensa esperada al realizar una acción. Entonces se la puede definir por la ecuación de Bellman.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha * \epsilon \qquad (2)$$

$$\epsilon = r_{t+1} + \gamma \max_{a} Q\,(s_{t+1}, a_t) - Q(s_t, a_t) \qquad (3)$$

De donde, α representa la tasa de aprendizaje $(\alpha \in [0,1])$, r representa la recompensa $\forall\, r \in R$ y γ es la penalización $(\gamma \in [0,1])$.

El aprendizaje en un entorno vehicular puede aumentar o disminuir según la región en la que se encuentre de acuerdo a su espacio de estados; el algoritmo QL para calcular la función de valor-acción es sometido al entrenamiento en el mapa, de tal manera que tanto el mapa como el valor de estado se van calculando de manera incremental de acuerdo al número de episodios al cual ha sido sometido, a medida que el agente va aprendiendo, los valores de la matriz Q se van modificando según su recompensa y experiencia previa, de forma que al final del estudio en dicha región estos valores representen la forma más eficiente de tomar una acción en el espacio de estado en el que se encuentre el agente.

III. FORMULACIÓN DEL PROBLEMA

Se ha establecido la simulación de un escenario vial, en el cual se optimizará un esquema dinámico de RSUs utilizando la comunicación V2I en la red VANET para brindar servicios mediante el despliegue apropiado de dichas infraestructuras. Para determinar la mejor ubicación de las RSU y poder asignar recursos dinámicamente en función de las limitaciones de cobertura, capacidad de las antenas y así minimizar la infraestructura de la red vehicular. En primer lugar, se modela matemáticamente la optimización de RSU's con la finalidad de obtener una cobertura efectiva de los vehículos en el escenario de prueba, para lo cual se define el conjunto de sitios candidatos $C = \{c_1, c_2, c_3, \dots, c_M\}$ donde la posición de cada sitio está definida por el par ordenado $(xc_i\ , yc_i)$ para él i-ésimo sitio candidato. Ahora se considera N usuarios o vehículos en el escenario, de tal manera que se establece el conjunto $U = \{u_1, u_2, u_3, \dots, u_N\}$ y de la misma forma se tiene la posición del j-ésimo vehículo dada por $(ux_j\ , uy_j)$. Se definen los valores binarios siguientes: $\alpha_{i,j} \in \{0,1\}$.

$$\alpha_{i,j} = \begin{cases} 1\ si \text{ el vehículo } j \text{ es cubierto por la antena } i \\ 0\ caso\ contrario \end{cases}$$

El valor $z_i \in \{0,1\}$ se plantea para el sitio candidato, de forma que:

$$z_i = \begin{cases} 1\ si \text{ la antena } i\ es\ activa \\ 0\ caso\ contrario \end{cases}$$

De manera similar para los usuarios se define la cantidad $y_j \in \{0,1\}$:

$$y_j = \begin{cases} 1 \text{ vehículo } j \text{ es cubierto por al menos una antena} \\ 0\ caso\ contrario \end{cases}$$

La relación para los sitios candidatos y la cobertura de usuarios es la siguiente:

$$y_j \leq \sum_{i=1}^{M} \alpha_{i,j}\, z_i \qquad (4)$$

La relación 4 es una restricción del modelo matemático la cual muestra que al no existir un sitio candidato que cubra el j-ésimo vehículo su valor es 0, de tal manera que no está cubierto. Es necesario indicar en el modelo una restricción que fije un porcentaje de usuarios a cubrir por lo tanto la ecuación 5 que se muestra a continuación permite colocar un 90% de vehículos cubiertos. El valor porcentual se establece de trabajos relacionados con la literatura VANET, con el cual se logra una optimización aceptable con este modelo ILP como se menciona en [8].

$$\frac{1}{N} \sum_{j=1}^{N} y_j \geq 0.9 \qquad (5)$$

Para finalizar con el modelo ILP se plantea la función objetivo para minimizar el número de RSU que sean activas, como se puede ver en la ecuación 6:

$$\min \sum_{i=1}^{M} z_i \qquad (6)$$

Una vez optimizada la infraestructura de la red VANET con el anterior modelo, ahora se define para la asignación dinámica de recursos (canales) modelar el algoritmo Q-Learning para el aprendizaje de la matriz Q donde se almacenan valores lógicos de acuerdo con el entrenamiento en el escenario de movilidad; la demanda y la capacidad total del controlador son parámetros interpretados por la tabla Q de estado-acción correspondiente al aprendizaje por refuerzo, para este aprendizaje al ser un MDP se utilizan las ecuaciones 1, 2 y 3. La función Q definida en la ecuación 2 es encontrada con el objetivo de almacenar la política óptima al seleccionar la acción que mayor recompensa entregue como se establece en la ecuación 1. Es decir, el agente selecciona de acuerdo con el estado en el que se encuentre la cantidad adecuada del conjunto $K = \{k_1, k_2, k_3, \dots, k_N\}$ de canales que asignará adecuadamente en función a la capacidad de este.

Se presenta a continuación el pseudocódigo del algoritmo Q-Learning bajo el cual se modeló el aprendizaje del controlador para cumplir con el último objetivo de este documento el cual es aplicar técnicas de RL en una red vehicular para el análisis de la asignación dinámica de recursos desde la infraestructura con relación a la demanda vehicular y sus restricciones de cobertura.

Inicio

1. **Inicializar** parámetros de aprendizaje: α, γ
2. **Para** cada(s_t, a_t) **hacer:**
   Inicializar matriz Q = 0;
   Fin Para
3. Seleccionar estado inicial aleatorio;
4. **Para** cada episodio **hacer:**
   #Conjunto de acciones que empiezan en un estado (s_t) y terminan en el estado objetivo (s_g).
5. **Mientras** $s_t \; ! = s_g$ **hacer**:
   - **a)** **Seleccionar** una acción (a_t) aleatoria para el estado actual (s_t).
   - **b)** **Ejecutar** esta posible acción (a_t) para pasar al siguiente estado (s_{t+1}).
   - **c)** **Obtener** el valor de $Q(s_t, a_t)$:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\,(r(s_t, a_t) + \gamma \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$$

   - **d)** **Encontrar** a_t en s_t con mayor valor de Q.
   - **e)** **Fijar** el $s_t \leftarrow s_{t+1}$

   Fin Mientras
   Fin Para

Fin

De esta manera es como se logra obtener la política óptima para la asignación de recursos en la red VANET.

IV. ANÁLISIS DE RESULTADOS

En la figura 5 se muestra el escenario de la red VANET, el mismo que se refiere a un sector altamente transitado al ser cercano a la UPS-Quito; este caso de estudio es optimizado por el modelo ILP en intervalos de tiempo en donde por ejemplo en el primer intervalo de las 35 RSU desplegadas en el escenario inicial se logra obtener la utilización de únicamente 23 antenas para dicho intervalo de tiempo, el estudio se realiza en varios episodios en los cuales el valor de las antenas activas y optimizadas es variable y con este valor se determina la cantidad de RSUs necesarias para cubrir al menos al 90% de vehículos en la red, los cuales tienen cobertura.

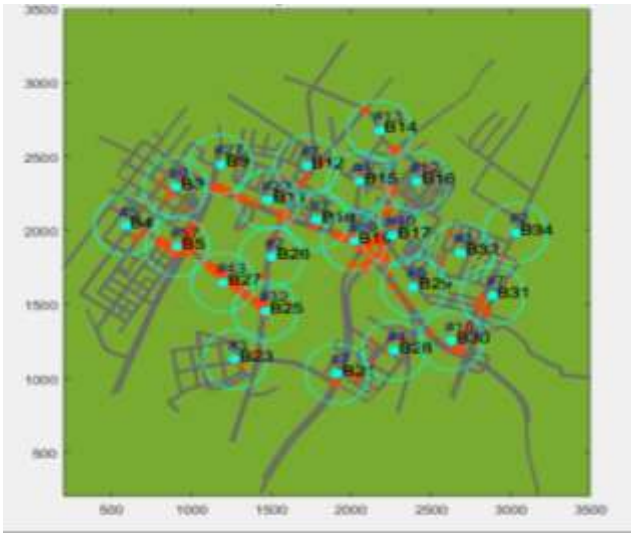

Figura 5: Escenario optimizado de la red VANET.

Se presenta ahora en la figura 6, el porcentaje de optimización de infraestructura en los diferentes intervalos de tiempo en los cuales se realizó el estudio; se puede observar que cuando existe una mayor cantidad de vehículos se utilizan hasta máximo el 80% de las RSU, es decir 28 antenas activas, logrando de esta manera optimizar el uso de la infraestructura de la red para la comunicación V2I.

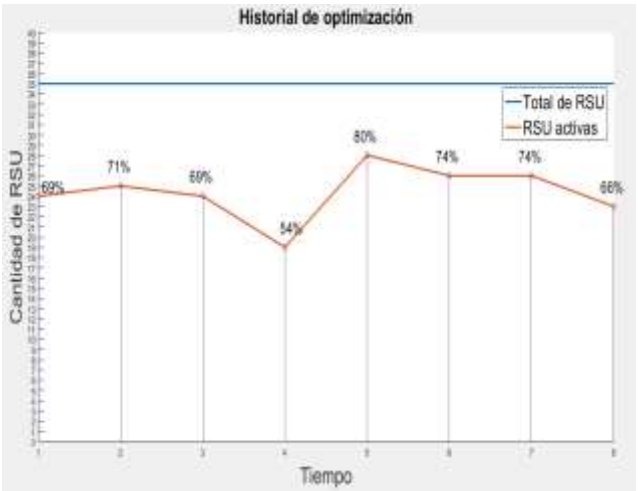

Figura 6: Historial de optimización en la red.

En la figura 7 se puede visualizar un cotejo de matrices Q, las cuales son las memorias del controlador de la red ad-hoc vehicular y contienen la información necesaria para la asignación dinámica de recursos después de un tiempo de

aprendizaje. Se evidencia que a mayor tiempo de entrenamiento se logra un mayor porcentaje de aprendizaje, con el cual se pueden tomar mejores decisiones para asignar de manera dinámica los canales necesarios a los usuarios y de esta manera asegurar la conexión. La matriz Q5 posee 85% de aprendizaje y es con la que se trabaja en este estudio, puesto que con este valor ya se puede hallar políticas óptimas.

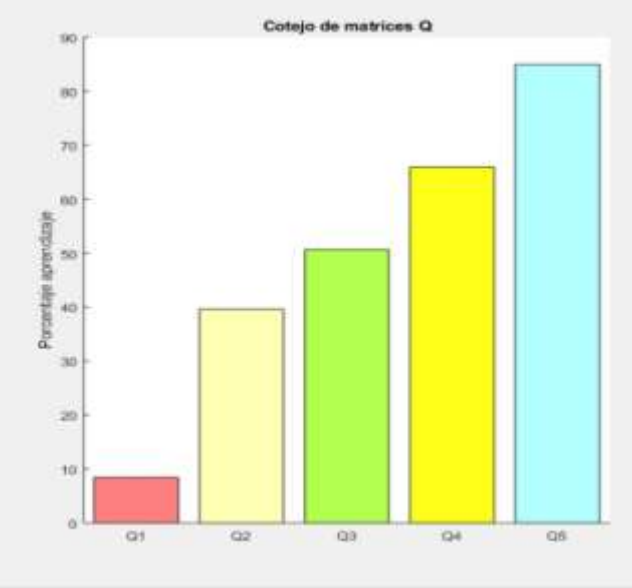

Figura 7: Cotejo de matrices Q

Una vez se tiene una matriz Q aceptable es necesario comprobar el comportamiento que tiene el controlador, el cual tiene un determinado número de canales para asignar a las RSU's activas, se puede observar en la figura 8 el comportamiento del controlador al terminar la ejecución del algoritmo, donde se aprecia que tiene menos canales disponibles, debido a que estos se asignaron en función a la demanda vehicular; la curva de la gráfica es estrictamente decreciente, ya que durante su funcionamiento y acorde a la demanda de vehículos el controlador va asignando canales, es decir, que si el tiempo tiende a infinito la capacidad del controlador tiende a cero, esto claro si todo el tiempo existe requerimiento de conexión en la red.

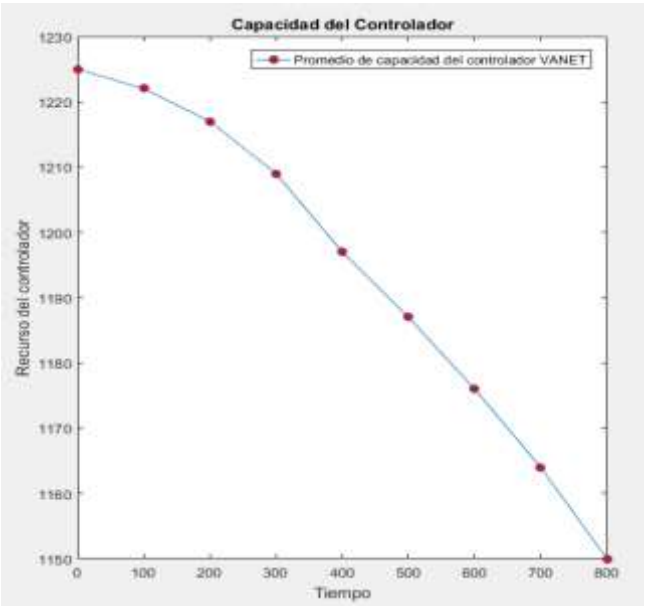

Figura 8: Comportamiento de controlador VANET

Por último en la figura 9 se tiene un historial de todo el comportamiento de la red al aplicar RL en donde la demanda vehicular está representada por la línea azul y la asignación de canales por la línea naranja y se logra visualizar que el controlador de la red asigna los canales necesarios y además se puede apreciar que en la RSU 5, se tiene una diferencia considerable de canales y usuarios lo cual representa un error de asignación con un valor de 16%, sin embargo, al analizar el comportamiento global de la red, es decir de las RSU activas durante todo el tiempo de estudio se tiene un porcentaje de acierto del 94% y un error en la asignación del 6%.

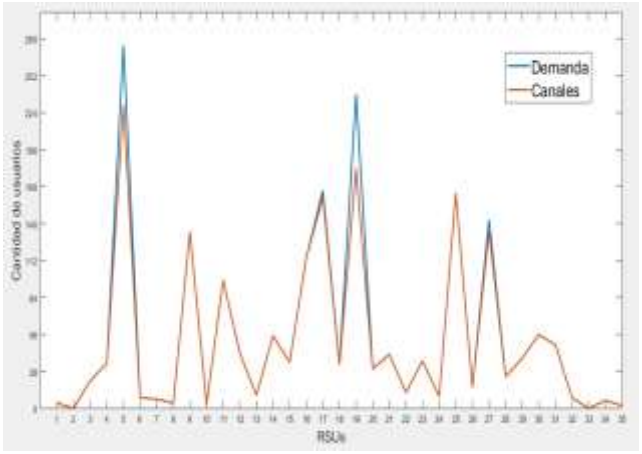

Figura 9: Historial de demanda vs canales

V. CONCLUSIONES Y FUTUROS TRABAJOS

En la investigación se estudia la asignación dinámica de recursos en un escenario de movilidad utilizando el algoritmo de Q-Learning para lo cual se logró entregar la distribución óptima de recursos con hasta un 6% de error en los escenarios con mayor demanda vehicular, basándose en la capacidad del controlador los resultados obtenidos muestran una técnica adecuada para las redes VANET. Además, se redujo el número de activación de las RSUs hasta en un 20% de un total de 35, optimizando el despliegue de la infraestructura basada en el tráfico vehicular en un determinado periodo de tiempo, es decir se activarán 28 antenas teniendo en cuenta que el área total de cobertura cumple con el 90% de autos conectados a la red. El algoritmo Q-Learning a través de su matriz de estado-acción consigue una respuesta deseada en 24 iteraciones en un promedio de 19200 segundos. Para discriminar que matriz Q se debe aplicar como memoria del controlador se evaluó mediante prueba y error en el escenario, llegando a tener un rendimiento promedio de la red del 94% de acierto y esto únicamente con un aprendizaje del 85% de la matriz Q.

En trabajos futuros se debe tener en cuenta los tiempos de entrenamiento del algoritmo QL que representa un alto consumo de recursos computacionales. Con los resultados obtenidos se puede pensar que esta técnica ayudará a descongestionar las redes inalámbricas de tal manera que se brinden los diferentes servicios que necesitan las redes ad-hoc vehiculares.

Referencias

[1] K. Mehta, L. G. Malik, and P. Bajaj, "VANET: Challenges, issues and solutions," *Int. Conf. Emerg. Trends Eng. Technol. ICETET*, pp. 78–79, 2013.

[2] J. C. Cuastumal and R. C. Simba, "Simulación de Escenarios de Comunicación Inalámbrica Entre Vehículos En El Sur De Quito Por Medio De Protocolos De Enrutamiento Basada En Tecnología Vanet," 2019.

[3] B. Patel, F. Khatiwala, and V. Reshamwala, "Traffic Information Verification Techniques in Vanet," pp. 551–553, 2017.

[4] J. Chi, Y. Jo, H. Park, and S. Park, "Intersection-priority based optimal RSU allocation for VANET," *Int. Conf. Ubiquitous Futur. Networks, ICUFN*, pp. 350–355, 2013.

[5] S. Paper, P. Santana, and R. C. Simba, "Optimizar Y Dimensionar La Ubicación De Los Rsu En Una Red Vial Mediante Modelamiento Matemático Basado En Ilp Para Determinar La Mejor Posición En La Infraestructura De Comunicaciones Vanet," 2019.

[6] S. Jiang, Z. Huang, and Y. Ji, "Adaptive UAV-Assisted Geographic Routing with Q-Learning in VANET," *IEEE Commun. Lett.*, vol. 7798, no. c, pp. 10–14, 2020.

[7] H. Faris and S. Yazid, "Development of communication technology on VANET with a combination of ad-hoc, cellular and GPS signals as a solution traffic problems," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, 2019.

[8] R. Cumbal, S. Gutiérrez, and C. Guerrero, "Asignacion Óptima De Recursos En Ambientes Móviles Dinámicos Desde La Infraestructura Vanet," pp. 1–5, 2019.

[9] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 1, pp. 375–388, 2021.

[10] N. C. Luong *et al.*, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[11] S. S. Doddalinganavar, P. V. Tergundi, and R. S. Patil, "Survey on Deep Reinforcement Learning Protocol in VANET," *1st IEEE Int. Conf. Adv. Inf. Technol. ICAIT 2019 - Proc.*, pp. 81–86, 2019.

[12] B. R. Kiran *et al.*, "Deep Reinforcement Learning for Autonomous Driving: A Survey," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–18, 2021.

[13] F. R. Yu and Y. He, *Deep Reinforcement Learning for Interference Alignment Wireless Networks*. 2019.

[14] S. Feki, A. Belghith, and F. Zarai, "A reinforcement learning-based radio resource management Algorithm for D2D-based V2V communication," *2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019*, pp. 1367–1372, 2019.

[15] L. Liang, H. Ye, and G. Y. Li, "Toward Intelligent Vehicular Networks: A Machine Learning Framework," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 124–135, 2019.

[16] Z. Hai-Tao, D. Ai-Qian, Z. Hong-Bo, L. Dapeng, and L. Nan-Jie, "Research on Q-Learning Based Channel Access Control Algorithm for Internet of Vehicles," *Proc. - 2016 Int. Comput. Symp. ICS 2016*, pp. 491–496, 2017.

[17] J. Wu, M. Fang, H. Li, and X. Li, "RSU-Assisted Traffic-Aware Routing Based on Reinforcement Learning for Urban Vanets," *IEEE Access*, vol. 8, pp. 5733–5748, 2020.

[18] R. Cumbal, C. Guerrero, and R. Hincapi, "Optimal Resources Allocation From VANET Infrastructures In Dynamic Mobile Environments," 2019.

[19] L. Liang, H. Peng, G. Y. Li, and X. (Sherman) Shen, "Vehicular communications: A physical layer perspective," *arXiv*, vol. 66, no. 12, pp. 10647–10659, 2017.

[20] W. Viriyasitavat, M. Boban, H. M. Tsai, and A. Vasilakos, "Vehicular communications: Survey and challenges of channel and propagation models," *IEEE Veh. Technol. Mag.*, vol. 10, no. 2, pp. 55–66, 2015.

[21] R. Cumbal, H. Palacios, and R. Hincapie, "Optimum deployment of RSU for efficient communications multi-hop from vehicle to infrastructure on VANET," *2016 IEEE Colomb. Conf. Commun. Comput. COLCOM 2016 - Conf. Proc.*, 2016.

[22] S. J. Russell *et al.*, *Artificial Intelligence A Modern Approach*. 1995.

[23] R. S. S. and A. G. Barto, "Reinforcement Learning: An Introduction," *Decis. Theory Model. Appl. Artif. Intell. Concepts Solut.*, pp. 63–80, 2015.

[24] C. J. C. H. Watkins, "Learning from delayed rewards Watkins." 1989.

[25] P. S. J. Barrios, "Comparación de técnicas de aprendizaje por refuerzo jugando a un videojuego de tenis.," 2019.

[26] A. M. Printista, M. L. Errecalde, and C. I. Montoya, "Una implementación paralela del algoritmo de Q-Learning basada en un esquema de comunicación con caché Resumen," 2000.

[27] G. R. Gay, P. Salomoni, and S. Mirri, "Technical Note Q-Learning," *Encycl. Internet Technol. Appl.*, vol. 292, pp. 179–184, 2007.